

- 1 -

POLYMORPHIC CAG REPEAT-CONTAINING GENE AND USES THEREOFBACKGROUND OF THE INVENTION5 (a) Field of the Invention

The invention relates to hGT1 gene, a polymorphic CAG repeat-containing gene and its uses thereof for the diagnosis, prognosis and treatment of psychiatric diseases, such as schizophrenia.

10 (b) Description of Prior Art

Schizophrenia is a chronic brain disorder characterized by a behavioral syndrome combining in various degrees hallucinations, delusions, social withdrawal, affective flattening, disorganized behavior and formal thought disorders. It affects up to 1% of the general population and results in a lower level of social and occupational functioning. Many recent studies indicate that schizophrenia may originate from neural cell disturbances occurring in the developing/maturing brain. 15 Genetic factors are known to play a major role in the etiology of this disorder as demonstrated by extensive family, twin and adoption studies. However, the quest for genes conferring susceptibility to schizophrenia has been difficult and has not yielded consistent findings using both association and linkage studies. It is thought that these difficulties are in part due to heterogeneity in etiology, both of genetic and non-genetic origins, resulting in a highly variable phenotype with respect to age at onset, symptom profile, course of 20 illness, response to medication, long term outcome and performance on neuropsychological tests.

One promising avenue to guide research in this search for genes increasing susceptibility to schizophrenia may be to distinguish patients on the basis of 25 therapeutic response to neuroleptics. Indeed, while most schizophrenic patients are improved by neuroleptic

medication, a substantial number of subjects (15 to 25%) remains severely symptomatic despite multiple and adequate neuroleptic therapeutic attempts. In contrast to this between subject variability, within subject (from one episode to the other) consistency of neuroleptic response have been reported. Clinical pre-treatment characteristics that correlate with good neuroleptic response include spontaneous high blink-rate and blink-rate decrease under Haloperidol challenge, absence of spontaneous movement disorders, and absence of dysphoric reaction within 24-48 hours of neuroleptic initiation. On a long term basis, it has been demonstrated that good response to neuroleptics (but not the severity of the symptoms prior to neuroleptic medication) in the early stages of the disease, predicts a better outcome. Neurophysiological characteristics that correlate with good neuroleptic response include high-frequency waves and few alpha and slow waves in computerized EEG prior to the treatment with neuroleptics, a specific profile of changes in quantified EEG spectrum under neuroleptics and high degree of electrodermal activity prior to neuroleptic treatment. An important number of studies indicate that dopamine neurotransmission is disturbed predominantly in the responsive schizophrenic patients. High pre-treatment plasma levels of HVA have been shown to predict good response to neuroleptics in most of the studies. Preliminary genetic epidemiological data indicate that poor or delayed response to neuroleptic treatment is associated with an increased prevalence of schizophrenia spectrum disorders in relatives of schizophrenic probands. These convergent lines of evidence suggest that long term response to neuroleptic medication may be considered as a bioclinical dimension with an etiologically relevant significance; the two extremes of

this dimension being occupied by two groups of schizophrenic patients, at least partially, distinct with respect to the pathogeny of their illness.

5 It would be highly desirable to be provided with a tool for the diagnosis, prognosis and treatment of psychiatric diseases, such as schizophrenia.

SUMMARY OF THE INVENTION

10 One aim of the present invention is to provide a tool for the diagnosis, prognosis and treatment of psychiatric diseases, such as schizophrenia.

15 Another aim of the present invention is to detect association between allelic variants of CAG repeat-containing genes and schizophrenia or its phenotypic variability with respect to long term response to neuroleptic medication.

20 In accordance with the present invention, we compared the allelic frequencies of various polymorphic candidate genes between two groups of schizophrenic patients carefully screened on the basis of their long term response to typical neuroleptics (excellent responders, Rs; non-responders, NRs) and controls. This report summarizes our finding while considering CAG containing genes as candidates for schizophrenia.

25 This family of candidate genes was deemed attractive for the following reasons: (1) CAG repeat instability was associated with several neurodegenerative brain diseases that display genetic anticipation, a feature believed to be present in schizophrenia, (2) some isolated, though promising, reports indicate that expanded CAG repeats are more prevalent in schizophrenic patients compared to normal controls, (3) CAG repeats are often very polymorphic and have been found to be over represented in coding sequences of the human genome particularly those coding for DNA-binding proteins/transcription factors. These factors are impor-

tant actors in the regulation of the genetic program and neurodevelopmental processes and have been implicated in several human neurodevelopmental diseases including one that may present with schizophrenia-like 5 symptoms, and; (4) CAG repeats (or the polyglutamine stretches for which they encode) might modulate the function of the genes (or protein) they are part of suggesting that they might be functional polymorphisms and not silent ones.

10 In accordance with the present invention there is provided a hGT1 gene containing transcribed polymorphic CAG repeat, which comprises a sequence as set forth in Fig. 3 and Figs. 4A-4C.

15 The allelic variants of CAG repeat of hGT1 gene may be associated with schizophrenia, affective diseases such as manic depression, neurodevelopmental brain diseases or with phenotypic variability with respect to long term response to neuroleptic medication.

20 More precisely, there are 5 allelic variants of CAG repeat which are identified as follows:

| Size of PCR amplified fragments (bp) | Predicted No. of CAG repeats | Shortest to longest |
|--|---------------------------------|------------------------|
| 171 | 11 | -3 |
| 174 | 12 | -2 |
| 177 | 13 | -1 |
| 180 | 14 | 0 |
| 183 | 15 | 1 |

25 In accordance with the present invention there is provided a method for the prognosis of severity of schizophrenia of a patient, which comprises the steps of:

- a) obtaining a nucleic acid sample of the patient; and
- b) determining allelic variants of CAG repeat of the hGT1 gene, and wherein long allelic variants are indicative of severe schizophrenia.

5 The preferred nucleic acid sample used in accordance with the present invention is DNA. For RNA sample, an additional step is carried out, which consists in using a reverse transcriptase to transcribe the RNA 10 into DNA.

10 More precisely, the allelic variants identified as short or as having between about 171 and 177 bp (referred to as -3, -2 and -1) are associated with mild schizophrenia and long or as having between about 180 15 and 183 bp (referred to as 0 and 1) are associated with severe schizophrenia.

15 In accordance with the present invention there is provided a method for the identification of patient responding to neuroleptic medication, which comprises 20 the steps of:

- a) obtaining a nucleic acid sample of the patient; and
- b) determining allelic variants of CAG repeat of the hGT1 gene, and wherein short allelic variants are indicative of neuroleptic response.

25 More precisely, the allelic variants identified as short or as having between about 171 and 177 bp (referred to as -3, -2 and -1) are associated with patient capable of neuroleptic response and long or as 30 having between about 180 and 183 bp (referred to as 0 and 1) are associated with non-response to neuroleptic medication.

35 In accordance with the present invention there is provided a non-human mammal model for the hGT1 gene, whose germ cells and somatic cells are modified to

express at least one allelic variant of the hGT1 gene and wherein the allelic variant of the hGT1 being introduced into the mammal, or an ancestor of the mammal, at an embryonic stage.

5 In accordance with the present invention there is provided a method for the identification of patient responding to neuroleptic medication, which comprises the steps of:

10 a) obtaining a nucleic acid sample of the patient;
and
b) determining allelic variants of CAG repeat of the hGT1 gene, and wherein short allelic variants (from about 171 to about 177 bp) are indicative of neuroleptic response.

15 In accordance with the present invention there is provided a method for the screening of therapeutic agents for the prevention and/or treatment of schizophrenia, which comprises the steps of:

20 a) administering said therapeutic agents to the non-human mammal of the present invention or schizophrenia patients; and
b) evaluating the prevention and/or treatment of development of schizophrenia in said mammal or said patients.

25 In accordance with the present invention there is provided a method to identify genes part of or interacting with a biochemical pathway affected by hGT1 gene, which comprises the steps of:

30 a) designing probes and/or primers using the hGT1 gene of the present invention and screening psychiatric patients samples with said probes and/or primers; and
b) evaluating the identified gene role in psychiatric patients.

In accordance with the present invention there is provided a method of stratifying psychiatric patients based on the allelic variants of the hGT1 gene for clinical trials purposes, which comprises:

- 5 a) obtaining a nucleic acid sample of the patients; and
- b) determining allelic variants of CAG repeat of the hGT1 gene, wherein patients are stratified with respect to their allelic variants and
- 10 wherein short allelic variants are indicative of neuroleptic response.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 illustrates the average allelic lengths of the GCT10D04 EST CAG repeat in controls, responsive (R) and non-responsive (NR) patients, showing the shorter (S) allele only, longer (L) allele only and the sum (L+S) of the two alleles in the three groups of subjects;

Fig. 2 illustrates the correlation between the average length of the (CAG)_n polymer of the short (a), the long (b) alleles and the sum of 2 alleles (C) and seventy of schizophrenia in the different classes of severity of the disease;

Fig. 3 illustrates the sequence homology between the human GCT10D04 sequence and the mouse GT1 gene; and

Figs. 4A-4C illustrate the nucleotide sequence of hGT1, wherein the upstream intron is in lowercase; Human gene sequence (exon) is in upper case; and the transcription start site ATG in bold.

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the present invention, the main objective was to detect allelic variants of CAG repeat containing genes associated with schizophrenia

or its phenotypic variability with respect to the presence or absence of schizophrenia and long term response to neuroleptic medication.

Accordingly, CAG repeat allelic variants were 5 compared between three groups of subjects: two groups of schizophrenic patients, one neuroleptic-responsive (Rs; n=43) and one -non-responsive (NRs; n=63), and; a group of controls screened out for DMS-IV axis I psychiatric disorders (C; n=87). Assessment of response 10 to conventional neuroleptics was based on a comprehensive review of medical files according to *a priori* defined criteria and blind to genotyping. Genes containing polymorphic CAG repeats were identified by means of genetic sequences data base searches.

15 The results in accordance with the present invention shows that short CAG repeat allelic variants of the hGT1 gene were associated with schizophrenia irrespective of neuroleptic response (% short alleles SCZ=45%; C=31%, p=0.005). This association was highly 20 significant in Rs (52%, p=0.0009) and marginal in NRs (40%, p=0.12) groups. A statistically significant correlation (Gamma=0.37, p=0.0024) between the CAG repeat length and the overall pattern of severity of schizophrenia was also observed.

25 Surprisingly and in accordance with the present invention, CAG repeat allelic variants of the hGT1 gene show strong association with neuroleptic responsive schizophrenia and length correlation with the overall pattern of severity of the disease.

30 The GT1 sequence includes a 5535 bp open-reading frame (ORF) of 5535 bps without interruption showing 85%homology to the mouse cDNA (Figs. 4A-4C). The sequence of GT1 is from one large (5276 bp) Bam HI fragment and three Pst I fragments (672, 200 and 371 35 bps). This ORF is preceded by a 490 bps intron

(including a consensus splice acceptor) and 19 bps of 5'-UTR. The entire ORF may be coded for by a single exon (we are still missing the sequences coding for the last 12 amino acids (36 bp)). While this type of genomic 5 organization is very peculiar and not often encountered several lines of evidence suggest that these sequences represent the GT1 gene. First, the presence of a splice acceptor upstream of the ORF suggest that the pre-mRNA will be processed. Second, the chromosomal 10 localization was determined by polymerase chain reaction (PCR) using the NIGMS somatic cell hybrid panel and two primers designed from our sequences. Sequencing of the previously described hGT1 alleles showed that they code for 10 to 14 glutamines (Q). The CAG-repeat 15 is generally constituted of 9 to 13 CAG repetitions followed by CAA (CAG9-13CAA) with the exception of the 13Q allele which is CAGCAACAG10CAA.

SubEI

Clinical

The study was conducted between 1994 and summer 20 1997. Patients have been recruited in the Douglas hospital (n=82), the Clinique Jeunes Adultes of L.H. Lafontaine Hospital (n=15) and the Schizophrenia Clinic of the Royal Ottawa Hospital (n=9). 333 schizophrenic patients were identified as potential subjects for this 25 study. 123 patients did not meet the criteria for schizophrenia or Rs/NRs (undifferentiated response) diagnoses. 125 and 85 patients met respectively the criteria for NRs-schizophrenia and Rs-schizophrenia. 62 NRs and 42 Rs subjects were not included in the study 30 because refusal or other exclusion criteria.

NRs schizophrenic patients were recruited according to the following criteria: (1) they all met axis I diagnosis of schizophrenia, according to the Diagnostic and Statistical Manual of Mental Disorders, 35 version IV (American Psychiatric association, *Diagnos-*

tic and Statistical Manual of Mental Disorders, APA; 1994) (DSM-IV), (2) they did not experience remission of psychotic symptoms within the past 2 years, (3) in the preceding 5 years, all patients underwent at least 5 3 periods of treatment with typical neuroleptics, from at least two distinct families of drugs, at therapeutic dosage (equal to or greater than 750 mg Chlorpromazine equivalent in monotherapy or 1000 mg chlorpromazine equivalent, when a combination of neuroleptics is 10 used), for a continuous period of at least 6 weeks at a time, with no significant relief of symptoms, and; (4) Unable to function without supervision in all or nearly 15 all domains of social and vocational activities with a Global Assessment Score (GAS) < 40 within the last 12 months.

Criteria for the selection of neuroleptic Rs patients were as follows: (1) all patients met the criteria for schizophrenia according to DSM-IV, (2) all 20 were admitted at least once to a psychiatric care facility because of acute psychotic episode, (3) during all hospitalizations, patients experienced full or partial remission in response to treatment with typical neuroleptics, at recommended dosage, within six-eight weeks of continuous treatment; remission being defined 25 as a rapid reduction of schizophrenic symptoms with limited residual symptoms, (4) all patients were able to function with only occasional supervision in all or nearly all domains of social and vocational activities with a GAS score ≥ 60 within the last 12 months, (5) no 30 patients had to be admitted to hospitals because of psychotic exacerbation, if and when compliant to treatment and treated continuously with typical neuroleptics, and; (6) at least one psychotic relapse when neuroleptic medication is reduced or interrupted. Exclusion 35 criteria for schizophrenic patients were brain

trauma, any neurological condition, drug or alcohol abuse in the last two years.

All schizophrenic patients were directly interviewed by the PI, a research psychiatrist trained in the use of the Diagnostic Interview for Genetic Studies (DIGS) (Nurnberger JI et al., *Archives of General Psychiatry*. 1994;51:849-59) and their medical records were comprehensively reviewed. Complementary information from the treating physician and nurses in charge of the patients and their close relatives was obtained, whenever possible. A best estimate diagnosis was established on the basis of all the available data. Responsiveness to typical neuroleptic medication was evaluated according to a 7 point's scale. The severity of symptoms and overall psychosocial functioning were assessed using the following instruments: (1) Brief Psychiatric Rating Scale (BPRS) (Woerner MG et al., *Psychopharmacology Bulletin*. 1988;24:112-117), (2) the Scale for the Assessment of Negative Symptoms (SANS), (3) Scale for the Assessment of Positive Symptoms, (4) the GAS, (5) the Pattern of Severity Scale, a 5 point's scale assessing overall course and outcome of the disease (American Psychiatric association, *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition, American Psychiatric association, Washington D.C.; 1994), and; (7) the Pattern of Symptoms subtypes, a categorical classification of patients according to the combination and changes over the course of the disease of positive and negative symptoms (Nurnberger JI et al., *Archives of General Psychiatry*. 1994;51:849-59). All these evaluation tools, except the BPRS are part of the DIGS.

The control group (C) was made by healthy volunteers recruited through advertisement in local papers (n=49) and married-in individuals from a linkage study

(n=38). All subjects in this group underwent a structured psychiatric interview in order to exclude those who meet DSM-IV axis I disorders. Subjects recruited through advertisement have also been screened for schizophrenia spectrum disorders and have been tightly matched for ethnic background (mother and father ethnicity) with schizophrenic patients. All, (except one responsive), patients and controls were Caucasians. All of them gave informed and written consent. The research protocol has been approved by the three hospitals ethic committee where the research took place.

Genetic methods

To identify sequences potentially encoding polymorphic polyglutamine tracts, we conducted a number of 15 Basic Local Alignment Search Tool (BLAST) (Altschul SF et al., *Journal of Molecular Biology.* 1990;215:403-410) searches using the following sequences: (1) (CAG)30 or (CAA)30 (BLASTn, unfiltered against the non-redundant nucleic acid and the expressed sequence Tag (dbEST) 20 databases) and, (2) Q30 (BLASTp, unfiltered, against the non-redundant protein database or tBLASTn against dbEST). Sequences containing homopolymer tracts of >7 CAG or CAA repeats or potentially encoding a tract of >12 glutamine residues were used to design PCR primers 25 able to amplify the CAG or CAA repeats. PCR primers were designed using DNASTAR Inc. (Madison, Wisconsin) software.

30 Genomic DNA was isolated from peripheral lymphocytes using standard methods. CAG repeat-containing fragments were amplified by PCR using specific primers for each repeat. PCR was performed in a total volume of 13 μ l containing 30 ng of human genomic DNA, 10mM Tris-HCl (pH 8.8), 1.5 mM MgCl₂, 50mM KCl, 1% Dimethyl-sulfoxide, 250 mM each of dCTP, dGTP, and dTTP, 25 mM 35 dATP, 1.5 uCi alpha 35S-dATP, 100 ng of each primer,

and 3 units of Taq polymerase (Perkin-Elmer). DNA was denatured at 94°C for 5 min., then subjected to 30 cycles of a 1 min. denaturation at 94°C, a 1 min. annealing at the optimized annealing temperature for 5 each primer pair and a 1 min. elongation at 72°C. This was followed by a final extension at 72°C for 5 min.

PCR products were electrophoresed on denaturing 6% Polyacrylamide gels and visualized by autoradiography. Absolute allele sizes were estimated according to an M13 sequence ladder. Since differences in absolute allele sizes were in all cases multiples of 3 base pairs, we assumed that variations in allele sizes were due to differences in the number of trinucleotide repeat units in the amplified sequences. By convention, we designated the most common allele as 0, with less common alleles as positive or negative integers according to their number of trinucleotide repeats (e.g. if allele 0 had 20 repeats, allele +2 and -2 would have respectively 22 and 18 repeats).

20 **Analysis:**

Each subject was assigned two numeric values which represent respectively the lengths of his short (S) and long (L) alleles. Under the assumption of a quantitative effect of the CAG tract length, data were 25 initially analyzed using a non-parametric analysis of variance (Kruskal-Wallis median statistic) where the independent variable is the diagnostic status (Rs, NRS and C) and the dependent variable is the length of the CAG repeat of S, L or the sum of the two alleles. In 30 the case of a significant overall group effect in the ANOVA, pair-wise contrasts between the different groups were performed using the Mann and Whitney U-statistic. This approach allows to control for the inflation of type I error secondary to multiple testing.

We also analyzed data by contrasting allelic frequencies in different pairs of groups using the χ^2 statistic. Alleles were grouped in different classes in accordance to the pattern of results found in the 5 analysis of variance. Since both patient and control groups include an important number of subjects with a French Canadian ethnic origin, we reanalyzed any association finding after stratifying subjects according to the ethnic origin of their parents (both parents from 10 French Canadian origin vs. at least one parent with non-French Canadian origin). This analysis allows to control for associations resulting from ethnically based differences in allelic frequencies (population stratification) as opposed to those attributable to the 15 pathological condition under study (true association).

When a particular EST showed allelic or size association with schizophrenia and/or responsiveness to medication, further analyses were performed to investigate the putative relation linking various clinical 20 dimensions (age at onset, pattern of severity) to the length of the CAG repeat. For this purpose, we used the Gamma correlation statistic, a non parametric statistic recommended when there are many ties in the data set. Clinical dimension that were used as criteria to define 25 the two groups of patients (GAS, severity of current symptoms, neuroleptic responsiveness scores) were not included in this analysis. Relations between categorical variables (schizophrenia subtypes of illness, pattern of symptoms) and the CAG alleles were explored by 30 a χ^2 statistic with the appropriate degree of freedom. Logistic regression was used to determine the attributable risk conferred by any EST allelic variants which showed a positive association with schizophrenia or 35 neuroleptic responsiveness. All analyses were made using the Statistica software (Statsoft).

Table 1. Demographic and clinical characteristics of patients and controls.

| | Non-responsive(62) | Responsive (43) | Controls (C) (87) |
|--|-------------------------|---------------------|-------------------|
| Mean age in years \pm SD (n) | 38 \pm 7 (62) | 40 \pm 10 (43) | 44 \pm 13 (87) |
| Education in years \pm SD (n) | 11 \pm 2.0 (59) | 11 \pm 2.8 (43) | 14 \pm 3.3 (49) |
| SES of HH \pm SD (n)* | 54 \pm 24 (53) | 51 \pm 24 (41) | 59 \pm 20 (49) |
| Sex, % M | 74% | 67% | 45% |
| Ethnic origin FC/OB | 27/35 | 26/17 | 40/47 |
| Subtype, U/P/D/C | 27/30/4/1 | 6/3 6/1/0 | — |
| Mean age at C ₁ in years \pm SD (n) | 18 \pm 3.9 (55)** | 24 \pm 4.8 (43) | — |
| Illness duration in years \pm SD (n) | 20 \pm 7.0 (55) | 16 \pm 8.8 (43) | — |
| % time as in-patient**(n) | 62% (61)** | 8.2% (43) | — |
| BPRS total score \pm SD (n) | 49 \pm 8.9 (53) ** | 24 \pm 3.9 (53) | — |
| NLP response score | 1.83 \pm 0.74 (58) ** | 6.3 \pm 0.67 (43) | — |
| Pattern of severity | 4.0 \pm 0.0 (55) ** | 1.9 \pm 0.5 (43) | — |

*SES of HH indicates socioeconomic status head of household; FC/OB, French Canadian/other ethnic background; U, undifferentiated; P, paranoid; D, disorganized; C, catatonic schizophrenia; C₁, first consultation; BPRS, Brief Psychiatric Rating Scale; and, NLP neuroleptic. **, p<0.001.

Table 1 shows the demographic and clinical characteristics of the three groups of subjects Rs, NRs, 5 and C. The two groups of patients were comparable with respect to age, level of education and socioeconomic status of the head of household. As expected, they differed significantly according to the severity of psychosis (BPRS scores, F=280, p<0.000), the percent of 10 time spent as inpatient since their first contact with the psychiatric institution (F=81, p<0.000) and the age at first contact with psychiatric care facilities (F=47, p<0.000).

Table 2 shows the sixteen different candidate 15 expressed CAG repeats identified and analyzed and includes mapping, homology, and polymorphism information.

Table 2: list of different studied ESTs

| Sequence ID | PCR | RT-PCR | PNQ | Homology information, potential function | Polymorphic | Map data |
|-------------|-----|----------|-----|---|-------------|----------|
| T08930 | + | na | 15 | homology with a human RNA-binding-protein CUG-BP/hNab50 and xenopus ETR-1 gene | - | - |
| R98242 | - | + | 27 | homology with a cAMP-responsive transcriptional activator regulating late gene expression | - | - |
| L37868 | + | na | 21 | N-Oct-3 TF, POU domain TF | - | HCH6 |
| U23868 | + | na | 26 | No known homology. | + | HCH1 |
| U23862os | + | na | 7 | possible homology with transporter-like protein (<i>S. cercisiae</i>) | - | - |
| N55395 | + | na | 15 | human zing finger protein TF | + | - |
| L10379 | + | na | 28 | no known homology | +++ | - |
| Z78314 | + | na | 20 | no known homology | ++++ | - |
| X85326 | - | not done | 11 | no known homology | na | - |
| T90581 | - | + | 10 | no known homology | - | - |
| L10375 | - | + | 16 | no known homology | + | - |
| X82209 | + | na | 27 | Q, human rml gene disrupted by a balanced translocation in meningioma | ++ | - |
| D26155 | + | na | 23 | SWI2/SNF2, a wide range transcription factor, interacts with ER and RA receptors | ++ | - |
| GCT5E11 | + | na | 22 | no known homology | +++ | HCH3 |
| TATA BF | + | na | 39 | TFIID TATA box binding protein, general transcription factor | ++++ | - |
| GCT10D04 | + | na | 14 | homology to a mouse retinoic acid inducible gene and stromelysin PDGF TF. | +++ | HCH17 |

PCR indicates Polychain Reaction; RT-PCR, reverse transcription PCR-reaction; PNQ, potential number of encoded polyglutamines, TF transcription factor; ER, estrogen receptor. RA, retinoic acid receptor.

Seven of the candidate sequences showed homology or identity with DNA binding domains or transcription factors. Most of the candidates (12/16) gave a PCR product with the predicted size. Candidates that amplified a larger than expected fragment or no products at all were further analyzed by RT-PCR to control for possible intronic interruptions in the genomic DNA. 5 Three candidates gave an RT-PCR product of the predicted size; only one was polymorphic using a small sample of chromosomes. Overall, 10/16 candidate sequences contained a polymorphic CAG repeat. Allelic frequencies of these polymorphic CAG repeats were compared in the four groups of subjects. 10 15

Only *allelic variants of the GCT10D04 locus (primers;* *SCZ15:GGGGCAGCGGGTCCAGAATCTTC,* *SCZ16:*

JKC7

WIC

TGGCCTTGCTGCCGTAGTGCT; annealing temperature 62°C) showed an overall significant group effect for the L allele (Kruskal-Wallis H (2, N = 194) = 12.18, p = .002), the CAG repeat average length being the shortest in the neuroleptic-responders (Rs), intermediate in the non-responders (NRs) and longest in the control group (C) (Fig. 1).

10 The reference point to measure the CAG repeat length is the most common allele (180 bp fragment or 14 predicted repeats), which is taken as 0. Alleles with n repeats above or below the 0 allele are scored +n or -n. C indicates the control group; Rs, neuroleptic responsive schizophrenic patients group; and NRs, neuroleptic non-responsive schizophrenic patients group.

15 A similar trend was observed for the S allele (Kruskal-Wallis H (2, N= 194) = 5.32, p =0.06). Post-hoc analysis using the U-statistic showed that this global effect was mainly due to the difference between neuroleptic-responders and normal controls (C) (L 20 allele: adjusted-Z=-3.52, p=0.0004; S allele: adjusted-Z= -2.28, p=0.02). Resistant schizophrenic patients showed also a trend toward smaller CAG repeat average size of the L allele compared to controls (C) (adjusted-Z=-1,68, p=0.09). When we analyzed the sum 25 of the two alleles, the three groups were statistically different (p=0.01) and the difference between controls and Rs was significant at the level of p=0.004 (adjusted-z=-2.8). Further analysis were carried out, testing the hypothesis that short alleles of the 30 GCT10D04 were more frequent in schizophrenic patients. For that purpose, two distinct classes of alleles, long (0,1) and short (-3, -2, -1), were defined and allelic frequencies between the four groups were reexamined (Table 3).

Table 3: frequencies of the CAG Allele short variants of hGT1 gene

| Schizophrenic patients | | | | |
|--|----------|---------|----------|--------|
| | Controls | SCZ | Rs | NRs |
| Number (2n) | 174 | 212 | 86 | 126 |
| % of short alleles | 31% | 45% | 52% | 40% |
| χ^2 | | P=0.005 | P=0.0009 | P=0.12 |
| Both parents are French Canadian | | | | |
| Number (2n) | 80 | 106 | 52 | 54 |
| % of short alleles | 35% | 47% | 54% | 41% |
| χ^2 | | P=0.09 | P=0.03 | P=0.5 |
| at least one parent is non French Canadian | | | | |
| Number (2n) | 94 | 106 | 34 | 72 |
| % of short alleles | 28% | 42% | 50% | 39% |
| χ^2 | | P=0.03 | P=0.018 | P=0.12 |

Allelic frequencies are given as percent of alleles shorter than 0 (<0). Frequencies are analyzed according to different diagnosis groups and ethnic background of parents. All frequencies were contrasted with the frequencies of the alleles shorter than 0 in the control group. SCZ indicates schizophrenic patients; Rs, neuroleptic-responsive schizophrenic patients; NRs, neuroleptic non-responsive schizophrenic patients, and; χ^2 , Chi 2 statistic with 1 degree of freedom

Schizophrenic patients, irrespective of their neuroleptic response status were more likely to carry one of the short alleles compared to controls ($\chi^2=7.6$, df=1, p=0.005). This difference was mainly due to Rs schizophrenic patients who were significantly more likely to have small alleles compared to controls ($\chi^2=11.0$, df=1; p=0.0009) and to NRs patients ($\chi^2=3.30$, df=1, p=0.07). Neuroleptic-non responders were marginally different from controls ($\chi^2=2.41$, df=1, p=0.12). When subjects with both parents of French Canadian origin or those with at least one parent from non French

Canadian origin were analyzed separately, the same pattern emerged (Rs vs. C: $\chi^2=4.6$, df=1, p=0.03; schizophrenics vs. C: $\chi^2=2.7$, df=1, p=0.09).

Finally, a correlation analysis indicated that 5 the size of the CAG repeat tract is linearly related to the pattern of severity of schizophrenia (measured blindly to genotype, using a 1-5 scoring system defined as follows: (1, *episodic shift*) episodes of illness interspersed between periods of health or near normality, (2, *mild deterioration*) periods of illness occur, but there are periods of return to near normality, with some ability to work at a job and near normal or normal social functioning, (3, *moderate deterioration*), the subject may occasionally experience some resolution of 10 symptoms, but overall the course is downhill culminating in a relatively sever degree of social and occupational incapacitation, (4, *severe deterioration*), the subject illness has become chronic resulting in inability to maintain employment (outside of a sheltered workshop) and social impairment, and; (5, *relatively stable*), the subject illness has not changed significantly (since it started at a severe level of impairment); the longer the size, the worse and poorer is the 15 outcome (Gamma statistic for S, L and L+S alleles 20 respectively: 0.25, p=0.01; 0.37, p= 0.002; 0.29, p=0.002) (Fig. 2).

To evaluate the proportion of variance attributable 25 to the CAG polymorphism in the phenotype responsive schizophrenia (as contrasted to the phenotype normal controls), we performed a logistic regression where the S and L alleles were the independent variables. This analysis indicate that the length of the two alleles contribute 10 % to the variance of this phenotype.

A sequence homology search was performed using the GCT10D04 nucleic acid sequence (GenBank acc. no. G09710) against the non-redundant nucleic acid database (BLASTn, GenBank). The GCT10D04 sequence was 84% homologous to a mouse gene (GT1, GenBank D29801, see figure 3) from which is transcribed a 7.2 kb cDNA encoding a 196 kDa protein of unknown function, suggesting that GCT10D04 represents a portion of the human homologue, which we term hGT1. The murine GT1 gene is inducible with retinoic acid in the mouse embryonic carcinoma cell line P19 and is expressed at highest levels in neurons but not in glial cells. A sequence homology search using the mGT1 protein sequence identified several conserved domains in another mouse gene (stromelysin PDGF responsive element binding protein transcription factor, GenBank U20282) and in its human homologue (AR1, GenBank U19345), suggesting that the hGT1 protein may also function as a transcription factor.

Common allelic variants, rather than rare mutations, may be responsible for the familial aggregation observed in complex diseases such as schizophrenia. Allelic variants that are neither necessary nor sufficient to cause a disease may not be identified by linkage analysis, particularly when the attributable risk is less than 10%. In contrast, association studies are sensitive to detect such variants.

To identify genes that may confer susceptibility to schizophrenia and/or its phenotypic variability with respect to neuroleptic responsiveness, we recruited patient according to their long term responsiveness to neuroleptic medication, a strategy that might reduce the putative genetic heterogeneity of schizophrenia. Control and patient groups were stratified according to

the ethnic background of parents; thus reducing the risk of population stratification bias.

In accordance with the present invention, neuroleptic-responsive-schizophrenic patients were significantly more likely to have hGT1 gene alleles with short CAG repeats as compared to patients who are characterize by long term poor response to neuroleptics and outcome. Furthermore, a significant correlation between the size of the hGT1 CAG repeat and the pattern of severity of the disease (the longer is the CAG repeat the more severe is the outcome) was identified in the group of schizophrenic patients regardless of the quality of their response to neuroleptic medication.

One major limitation of association studies with a relatively small number of subjects and a potentially high number of genes to be tested is an increased risk of false positive findings (type I error). In this study, we focused on candidate genes containing expressed and polymorphic CAG repeats, thus markedly reducing the number of genes to be tested; the number of CAG repeats is thought to be around 700 in the total human genome. Polymorphic CAG repeats containing transcripts might be much less represented. Based on these numbers, the Bonferroni corrected p-value for our tested hypothesis ought to be between 2×10^{-4} and 7×10^{-5} . In our study, and in spite of the small sample sizes, short alleles were likely to be more frequent in responsive schizophrenia compared to controls at a p-value of 9×10^{-4} , which is suggestive of a true association in the case of a complex disease such as schizophrenia. Moreover, the fact that the association is detected in an ethnically very homogenous subgroup (both parents French Canadians) as well as in a mixed subgroup (at least one parent is non French Canadian),

CROSS-REFERENCED DOCUMENT

suggests that this allelic association is very likely not to be due to stratified population bias. Furthermore, the fact that hGT1 gene has a high homology with a mouse gene involved in neural cell differentiation induced by retinoic acid is consistent with both the neurodevelopmental and retinoic acid hypotheses of schizophrenia.

Patients who presented with episodic shifts and good between-episode recovery were more likely to have shorter CAG repeats in both of their hGT1 alleles. This finding could be interpreted in various different ways: (1) it could indicate that hGT1 short alleles have a causative effect in the disease of patients with favorable outcome (good between episodes recovery, slow progression of functional deficits) whereas resistant patients with sever pattern of severity (continuos psychosis, no psychotic free episodes, rapid decline of psychosocial functioning) have other genetic or environmental factors involved in their disease. Patients falling between these two levels of severity may be a more mixed group difficult to relate to either one of the two extremes using clinical criteria (heterogeneity hypothesis), (2) it could indicate that the hGT1 polymorphism modulates the pattern of severity of the schizophrenia phenotype but not the susceptibility to schizophrenia *per se* (modifier gene hypothesis), and (3); hGT1 gene could influence susceptibility to schizophrenia irrespective of the pattern of severity and responsiveness to neuroleptics; the weak association in the group of resistant schizophrenic patients being the result of a selection bias. Indeed, should another gene with a higher attributable risk than the hGT1 be acting in the resistant form, the enrichment of hGT1 short alleles in non-responsive patients with severe pattern of the disease would be relaxed and the

association would be more difficult to identify in this group. In accordance with this hypothesis, family studies have suggested that neuroleptic-delayed response and marked deterioration in the psychosocial functioning are associated with a higher degree of familial aggregation of the disease; suggesting the presence of gene(s) with relatively high penetrance.

Transcription factors are major actors in all neurodevelopmental phases, and might be very important when developmental activity is intensive such as brain fetal development or synaptic pruning occurring in the adolescence phase of human development. They have been implicated in animal complex behavioral traits and have also a major role in the transduction pathways involved in the biological adaptation of the central nervous system to environmental changes (ranging from physical conditions such as viral infections to psychological conditions such as nurturing behavior in mice). It is also of interest to note that all antipsychotic drugs modulate DNA transcription in specific areas of the brain and ultimately results in modifications of neuronal interconnectivity. Variable number of tandem repeats, including trinucleotide repeats, have been found to be over represented in genes coding for DNA-binding proteins/transcription factors. Such repeats may be the basis of a fine modulation of gene activity. We speculate that one or multiple transcription factors might be involved in the etiology of schizophrenia or its phenotypic variability (including the quality of the response to different drugs). It is therefore of interest to consider transcription factors containing polymorphic CAG repeats as a putative candidate "family of genes" for schizophrenia and other psychiatric disorders thought to be of a neurodevelopmental origin.

The present invention will be more readily understood by referring to the following examples which are given to illustrate the invention rather than to limit its scope.

5 While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following,
10 in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set
15 forth, and as follows in the scope of the appended claims.